

BI Spektrum

EINE PUBLIKATION DES TDWI E.V.

Sonderdruck aus
BI-SPEKTRUM 4/2018

Self-Service-BI: Das Geheimnis des Erfolgs liegt in ausgewogener Governance ab Seite 8

Best Practice

Maschinelles Lernen bei der Bundesagentur für Arbeit

Seite 24

Fachartikel

Klare Strategie für DSGVO-konforme Datenplattformen

Seite 36

Interview

„KI ist der Motor der Digitalisierung“

Seite 20

Dr. Stefan Wess,
Empolis



Paradigmenwechsel: Self-Service-BI im Wandel

Self-Service-BI – mehr als eine Softwareauswahl

Ein Beitrag von
Martin Braun

Das Bedürfnis nach autonomer Informationsbeschaffung, unabhängig von IT-Abteilung und Controlling, existiert seitens Fachabteilung und Management schon, seitdem Business Intelligence in Unternehmen Einzug gehalten hat.

Sonderdruck aus
BI-SPEKTRUM 4/2018

Ursprünglich waren BI-Anwendungen als Desktop-Applikationen lokal beim Anwender installiert. Dieser war in der Lage, eigene Berichte zu entwerfen oder auch neue Datenquellen zu erschließen. Er agierte weitestgehend unabhängig von IT und Controlling. Mit dem Wechsel auf zentrale, servergestützte BI-Applikationen und Data Warehouses ging auch die Hoheit über Datenquellen und -modelle in die IT-Abteilung über. Berichte wurden meist ebenfalls zentral erstellt und gewartet. Der Verlust an Autonomie war dem Anwender nur schwer zu vermitteln. Sukzessive

hielten daher Analyse- und Query-Funktionalitäten wieder Einzug. Die Datenbereitstellung und -modellierung verblieben jedoch in der IT-Abteilung. Jede zusätzliche Anreicherung oder Änderung der Daten musste bei der IT-Abteilung beauftragt werden.

Diese Vorgehensweise entspricht jedoch nicht mehr den aktuellen Anforderungen einer agilen BI und dem Selbstverständnis ihrer Anwender. Der erforderliche Grad an Autonomie hinsichtlich Selbstbedienung in Daten und Auswertungen wird mit der Einführung von Self-Service-BI erreicht [ALS16].

Bild: Shutterstock



War vor wenigen Jahren der Bedarf an Ad-hoc-Auswertungen noch durch die Bereitstellung von ein paar Data Marts zu befriedigen, stellen die Self-Service-Anforderungen der bimodalen IT-Welt den CIO inzwischen vor große Herausforderungen. BI-Anbieter suggerieren am Markt, dass diese durch den Erwerb des „richtigen Werkzeugs“ bereits bezwungen wären. Doch das wäre deutlich zu kurz gedacht. Agilität und Heterogenität der Anforderungen, Datenvolumen und -aufkommen sowie Rahmenbedingungen durch Berechtigungskonzepte und Datenschutzauflagen erfordern ein Überdenken der BI-Architektur, -Infrastruktur und -Organisation.

Anforderungen an eine moderne bimodale BI

Eine häufig anzutreffende Definition beschreibt Self-Service-BI wie folgt [Gar18]: Self-Service-BI soll den Fachanwender in die Lage versetzen, eigene Reports, Analysen und Daten-Explorationen zu erstellen und mit anderen Anwendern zu teilen. Dies soll weitestgehend unabhängig von der IT-Abteilung möglich sein. Das Anreichern von lokalen Daten in dem bereitgestellten Datenmodell sollte ebenfalls möglich sein. Dies alles erfolgt in einer durch die IT-Abteilung validierten und unterstützten Architektur und Softwareumgebung.

Aus Sicht des Nutzers ergibt sich hieraus die Anforderung nach einem intuitiv bedienbaren Tool-Set, mit dem er seine Aufgaben eigenständig und schnell lösen kann, und einem Datenpool, der ihm die benötigten Daten in einer für ihn verständlichen Struktur und in der benötigten Aktualität bereithält (siehe Abbildung 1).

IT-Abteilungen fällt es zunehmend schwerer, mit den wachsenden Bedürfnissen im Wandel der digitalen Transformation Schritt zu halten. Fachabteilungen benötigen die kurzfristige Bereitstellung teilweise volatiler Anforderungen, für ein schnelles „Time to Market“ ihrer agilen BI-Projekte. Dies wird durch standardisierte IT-Prozesse ausgebremst. Für die IT-Abteilung geht Self-Service-BI unter anderem einher mit der Verarbeitung großer Datenvolumen, einer geeigneten Architektur und Struktur für deren Bereitstellung, der Einhaltung von Zeitfenstern für deren Aufbereitung, der Wahrung von Datenschutz- und Berechtigungsrichtlinien, der Gewährleistung der Datenintegrität und Schaffung eines architektonischen Rahmens, der korrekte Ergebnisse liefert, und der Vermeidung von Schattendatenbeständen.

Es bedarf einer BI-Landschaft der „zwei Geschwindigkeiten“: auf der einen Seite ein klassisches, verlässliches Standard-Reporting, auf der anderen ein experimentelles System, in dem innovative Lösungen geschaffen und auch wieder umgestoßen werden können. Dies bedeutet aber nicht, dass in einer bimodalen BI, an standardisierten IT-Verfahren vorbei, eine parallele Umgebung bereitgestellt wird, sondern vielmehr in der bestehenden Umgebung ein ausreichend weit gefasster Rahmen zu setzen ist, in dem sich eine Self-Service-BI entfalten kann.

MARTIN BRAUN berät mit über 15-jähriger Projekterfahrung die Kunden des IT-Beratungs- und Systemintegrationsunternehmens msg im Umfeld von Enterprise-Business-Intelligence-Infrastruktur von der Strategie und Konzeption über die Umsetzung bis zum laufenden Betrieb.

E-Mail: Martin.Braun@msg.group



Zielgruppen optimal bedienen

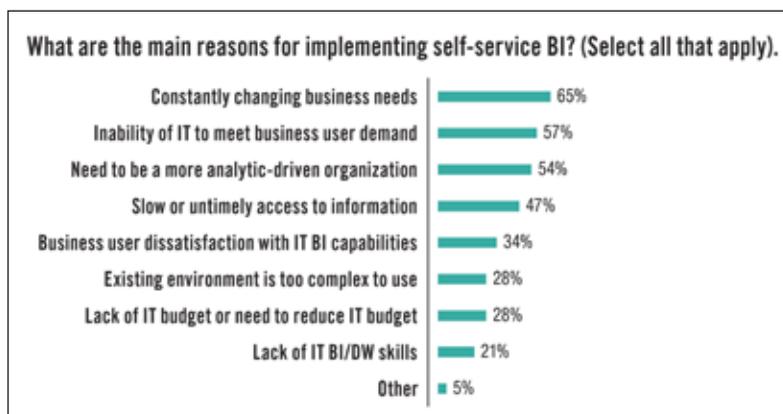
Die offerierte Datenstruktur muss allen Anwendergruppen im Unternehmen gerecht werden. Die Anwendergruppen lassen sich in drei Grundkategorien unterteilen [BAR18], die jeweils für sich den Begriff „Self-Service-BI“ unterschiedlich interpretieren (siehe Abbildung 2).

Das **Management**, das in der Regel nur eine geringe IT-Affinität aufweist, wünscht hochaggregierte KPIs und Kennzahlen, derer es sich nach Bedarf bedienen kann. Es erwartet ein Datenmodell, das schnell, valide und flexibel kombinierbar Informationen liefert. Eine zusätzliche Herausforderung stellt das Einbinden externer Daten dar, die sich bestenfalls intuitiv in das bereitgestellte Modell einfügen lassen. Hierfür eignen sich kleine Data Marts, in denen eine reduzierte Anzahl an Entitäten und Attributen, dafür aber eine große Zahl an vordefinierten Steuerungskennzahlen bereitgestellt werden.

Der **Fachbereich** erwartet detaillierte Prozessinformationen und möchte diese aus verschiedensten Blickwinkeln betrachten oder sich mit Data Discovery auf Entdeckungsreise begeben. Eine klassische Star-Schema-Architektur ist hierfür meist zu sperrig und unflexibel. Hier bietet sich eher ein Entitäten-orientiertes Snowflake-Schema an, das Geschäftsprozesse bereits valide abbildet, dem Anwender aber dennoch genügend Spielraum lässt, auch mit geringen IT-Kenntnissen alle erforderlichen Analysen zu tätigen. Der Anwender muss in der Lage sein, eigene Entitäten-Relations-Modelle zu erstellen. Das bereitgestellte Basisdatenmodell sollte hierbei mit einer homogenen Nomenklatur unterstützen.

Der **Data Scientist**, der hingegen über eine in der Regel gute Datenbank und Modellkenntnisse verfügt, erwartet maximale Autonomie. Er benötigt

Abb. 1: Gründe für die Implementierung von Self-Service-BI (Quelle: [ImW11])



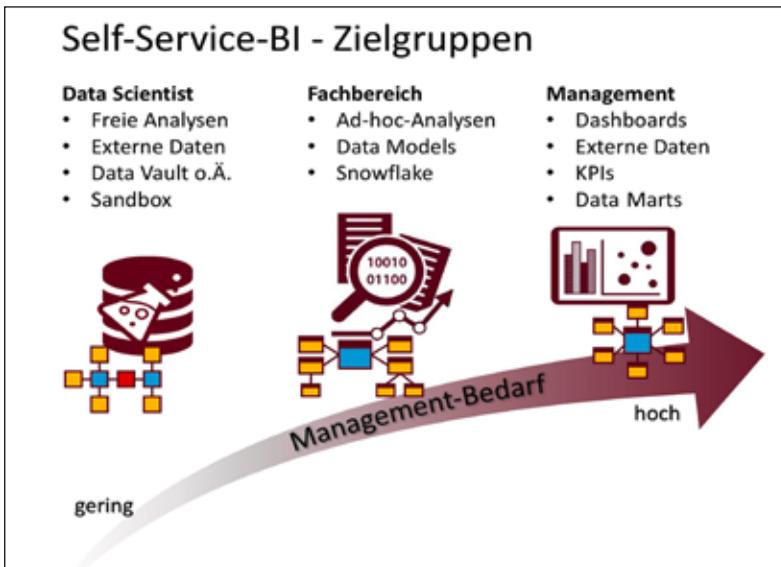


Abb. 2: Zielgruppen für Self-Service-BI

ein attributreiches Datenmodell, dessen Entitäten er nach eigenen Bedürfnissen verketteten kann. Starre Relationen wären hier fehl am Platz. Star- und Snowflake-Modelle sind hierfür eher ungeeignet. Ein Data Vault und eine performante Sandbox-Umgebung sollten diese Bedürfnisse optimal zufriedenstellen. Es sollte ebenfalls in Betracht gezogen werden, dass hierbei große Datenmengen auf sehr komplexe Abfragen treffen können. Daher sollte abgewogen werden, ob es nicht besser ist, für den Data Scientist eine Spielwiese, entkoppelt vom produktiven BI-System, bereitzustellen.

Security effizient implementieren

Self-Service-BI bedeutet nicht, dass jeder Anwender sich aller Daten beliebig bedienen kann. Es gilt, neben den unternehmensinternen Datenschutzrichtlinien auch die externen Anforderungen audit-sicher auszurollen. Letztere haben sich erst kürzlich durch die DSGVO noch einmal verschärft. Ein geeignetes Berechtigungskonzept muss sicherstellen, dass diese auf allen Architekturebenen für jeden Anwender greifen. Hierbei sollte im Fokus stehen, die Komplexität auf ein notwendiges Minimum zu reduzieren, um Fehler bei der Implementierung und Pflege zu verhindern.

Ob die Security-Filter auf Datenbankebene oder besser innerhalb des BI-Frameworks implementiert werden, hängt von der eingesetzten BI-Hard- und -Software ab. Es bietet sich jedoch an, Access Control Lists (ACL) zentral zu pflegen und diese in alle Bereiche einheitlich auszurollen. Wenn möglich, sollte darauf geachtet werden, dass Berechtigungen von den unteren Architekturebenen an die höheren automatisch vererbt werden.

Architektur und Infrastruktur - Schlüssel zum Erfolg

Hinsichtlich der Auswahl der geeigneten Infrastruktur und Architektur gilt es, eine Vielzahl von Kriterien zu beachten, von denen im Folgenden einige wichtige beleuchtet werden.

Zuerst sollten geklärt sein, welche **Anwendergruppen** und **Detaillierungsebenen** bedient werden müssen. Ist beispielsweise ein Data Vault notwendig oder gegebenenfalls aus Sicht der Datenversorgung sogar sinnvoll? Jedenfalls empfiehlt es sich, eher kurz- und mittelfristige Bedürfnisse bei der Entscheidungsfindung einfließen zu lassen. Langfristige Anforderungen können bis zum Eintreten des konkreten Bedarfs von technischen oder architektonischen Änderungen überholt worden sein. Zusätzliche Architektur-Layer bedeuten in der Regel auch höhere Kosten und ein größerer Aufwand an Hardware und Pflege.

Einen weiteren entscheidenden Punkt stellt das **Datenwachstum** und damit die erforderliche Aktualität dar. Welche Mittel sind erforderlich, um das Datenwachstum im Rahmen der zur Verfügung stehenden Zeit durch alle Ebenen zu bewegen? Sind für bestimmte Anforderungen Daten in Realtime oder Near-time erforderlich? Es sollte in Erwägung gezogen werden, höhere Layer anstatt physisch zu persistieren gegebenenfalls zu virtualisieren oder gänzlich auf Data-Virtualisation zu setzen. Dies verringert auch bei der Implementierung der Data Security die Komplexität, sofern eine benutzerbasierte Authentifizierung bis zur physischen Schicht realisiert werden kann.

Bei der **Data Security** steht man gegebenenfalls vor der Entscheidung zwischen einer Datenbank oder einer BI-applikationsseitigen Umsetzung, in Abhängigkeit von den technischen Gegebenheiten. Hat man die Auswahl zwischen beiden, sollte neben Performance und Komplexität der Lösung auch berücksichtigt werden, wer für die Pflege des jeweiligen Systems verantwortlich ist.

Beim Entwurf des Datenmodells sollte auf eine **einheitliche Nomenklatur** geachtet werden. Dies unterstützt den Anwender bei der Erstellung inhaltlich valider Berichte und Analysen. Zusätzlich empfiehlt sich die Bereitstellung eines Data Glossary, das unter anderem Auskunft über die Definition von Attributfeldern und die Berechnungslogik von Kennzahlen liefert. Dies vermeidet Fehlinterpretationen und fördert ein unternehmensweit einheitliches Verständnis der Daten.

Ist es gewünscht, dass Anwender lokale Daten wie zum Beispiel Vergleichsdaten von Marktbegleitern mit zusätzlichen Attributen anreichern dürfen, sollte berücksichtigt werden, dass sich in ihrem Datenmodell auch die zur Verknüpfung benötigten natürlichen Schlüssel wiederfinden und dem Anwender zur Verfügung gestellt werden. Die „natürlichen Schlüssel“ ermöglichen es dem Anwender überhaupt erst, externe Daten korrekt an den bestehenden Datenhaushalt anzuknüpfen. Häufig werden natürliche Schlüssel im Zuge der Datenaufbereitung gerade in Faktentabellen durch Surrogat-Schlüssel ersetzt oder dem Anwender bewusst ausgeblendet.

Bei der Auswahl der geeigneten Datenbank-Subsysteme sollten das erwartete **Datenvolumen** und die Erwartungshaltung hinsichtlich Abfrage-Performance neben Lizenz- und Wartungskosten in die Entscheidung miteinfließen. Nach wie vor werden

viele BI-Systeme auf zeilenorientierten Datenbanken implementiert, während eine In-Memory- oder spaltenorientierte Datenbank einen immensen Geschwindigkeitszuwachs und somit eine weitaus höhere Nutzerakzeptanz erreichen würde. Alle namhaften Anbieter von Datenbank-Software oder -Appliances bieten eine entsprechende Lösung im Portfolio an.

Sofern Sandbox-Datenbereiche Bestandteil des Lösungskonzeptes sind, sollte eine Entkoppelung auf separater Hardware oder mindestens als eigenständige Instanz konfigurierbar sein, will man nicht Gefahr laufen, dass eifrige Data Scientists die BI-Applikation lahmlegen.

Kriterien für die Softwareauswahl

Die Auswahl an Software, die Self-Service-BI anbietet, ist vielfältig. Bei der Entscheidung gilt es einige Kriterien zu beachten.

Passt die Applikation in die bestehende Systemlandschaft und erfordert sie nur einen geringen Pflegeaufwand seitens der IT-Abteilung? – Die durch die Autonomie der Anwender gewonnene Entlastung der IT-Abteilung kann durch die Pflege einer zusätzlichen Applikation schnell wieder kompensiert sein.

Wird die Software eine große Nutzerakzeptanz erreichen? – Ist die Applikation für den Nutzer nicht intuitiv bedienbar oder bestehen Defizite im Funktionsumfang, kann dies schnell das gesamte Self-Service-BI-Projekt in Frage stellen.

Bietet die Software BI-Funktionen, die eine Zusammenarbeit in Teams unterstützen? – Das Einsparpotenzial einer Self-Service-BI-Lösung kann nur dann optimal ausgeschöpft werden, wenn Nutzer Ergebnisse und Daten mit ihrem Team teilen können.

Eine geeignete Self-Service-BI weist mindestens folgende Funktionsmerkmale auf:

- Intuitive Erstellung und Modifikation von Ad-hoc-Reports, Analysen und Dashboards sowie Visualisierungen
- IT-unabhängige Integration lokaler privater Daten wie zum Beispiel .csv- und Excel-Dateien bis hin zu externen Datenquellen und Datenbank-Subsystemen
- Erstellung und Modifikation eigener Datenmodelle in einer Sandbox-Umgebung

Häufig werden auch Funktionsmerkmale wie Data Discovery, Microsoft-Office-Integration oder Unterstützung mobiler Endgeräte erwartet. Dies sind aber auch Anforderungen in klassischen BI-Umgebungen.

Für den nachhaltigen Erfolg einer Self-Service-BI-Strategie ist die Entscheidung für eine geeignete Gesamtarchitektur unabdingbar. Diese sollte als zentrales Element die Bedeutung der User-Experience berücksichtigen und somit zu einem großen Zuspruch und Nutzungsgrad beitragen. Die Auswahl der geeigneten Software ist zwar ein wichtiger Bestandteil für den Erfolg der Self-Service-BI, stellt dabei aber nur einen Teil des gesamten Puzzles dar.

Literatur

- [AIS16] Alpar, P. / Schulz, M.: Self-Service Business Intelligence. 2016, <https://www.semanticscholar.org/paper/Self-Service-Business-Intelligence-Alpar-Schulz/ae8e00f98427eed226b2c75a0ee562f2efd32e3>, abgerufen am 14.9.2018
- [BAR18] BARC: Self-Service BI: An Overview., abgerufen am 14.9.2018
- [Gar18] Gartner: IT-Glossary, self-service business intelligence. <https://www.gartner.com/it-glossary/self-service-business-intelligence>, abgerufen am 14.9.2018
- [ImWXX] Imhoff, C. / White, C.: Self-Service Business Intelligence Empowering Users to Generate Insights. TDWI Best Practices Report 2011